Confidence-Based Performance Assessments for the BMDO Family of Systems

Barry L. Mitchell
Thomas P. Spriesterbach

The Johns Hopkins University
Applied Physics Laboratory
Laurel, Maryland 20723

ABSTRACT

How models will support the T&E of ballistic missile defense systems is currently a topic of much debate. The authors have developed a methodology to extend weapon system test results to the theater-level using the Extended Air Defense Simulation (EADSIM). The primary Measure of Effectiveness (MOE) is protection effectiveness for the BMDO family of systems. EADSIM is run stochastically using offline random draws from a database of accuracy, timeline, and reliability distributions to quantify confidence in family of systems performance. Sensitivity analysis is conducted to identify performance drivers. This paper discusses details of the approach and presents results from a proof-of-principle test of the methodology.

INTRODUCTION

The primary operational requirement for a ballistic missile defense system is protection effectiveness, defined as the ratio of threat missiles killed divided by the total number of incoming threat missiles. Unfortunately this requirement is not directly testable for the Ballistic Missile Defense Organization (BMDO) Family of Systems (FoS) because test limitations restrict weapons testing to one-on-one and few-on-few situations. System evaluators will use force-on-force models to project FoS performance at the theater-level.

The current BMDO force-on-force models have limitations that prevent testers and system evaluators from assessing confidence in FoS performance. The simulations do not model the proper reliability, accuracy, and timeline distributions necessary for stochastic analysis of FoS protection effectiveness. Because of these limitations, much of the FoS analysis done by BMDO to date has been expected value analysis, that is, the analysis has not included performance variations due to random variations in the real world and uncertainties in weapon system performance.

We have developed a methodology to overcome these limitations by drawing random variates from the proper distributions and evaluating confidence in FoS protection effectiveness with Monte Carlo trials.

In this paper we discuss details of the approach and present results from a proof-of-principle test of the methodology. The results of the test are rather astounding and are a reminder that random variations in the real world can cause the FoS to perform radically different than predicted by expected value solutions. We also discuss sensitivity analysis using the proof-of-principle data to help identify performance drivers.

We used the Extended Air Defense Simulation (EADSIM) as the force-on-force simulation for the proof-of-principle test, but the methodology could also be applied using the Extended Air Defense Testbed (EADTB). Hopefully, lessons learned from this experiment will help guide future requirements for both models.

METHODOLOGY OVERVIEW

The problem with the current force-on-force models is they do not model the proper distributions for reliability, accuracy, and timelines. EADSIM, for example, draws random variates only from uniform distributions, not from Binomial, Gaussian, Lognormal, and Exponential Distributions as needed to properly characterize FoS performance. The situation with EADTB is not much better. Although the EADTB framework does offer a variety of probability distributions, the current set of EADTB Specific System Representations (SSRs) do not use them to suit the needs of testers and system evaluators doing stochastic analysis.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
27MAR2001	N/A	27MAR2001 - 29MAR2001

Title and Subtitle Confidence-Based Performance Assessments for the BMDO Family of Systems	**Contract Number**
	Grant Number
	Program Element Number
Author(s) Mitchell, Barry L.; Spriesterbach, Thomas P.	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) The Johns Hopkins University Applied Physics Laboratory Laurel, MD 20723	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) OSD Pentagon Washington, DC	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes Papers from the Proceedings AIAA 2nd Biennial National Forum on Weapon System Effectiveness, held at the John Hopkins University/Applied Physics Laboratory, 27-29 March 2001. Controlling Agency is OSD, Pentagon, Washington DC, per Richard Keith, Editor. See also ADM201408, non-print version (whole conference). , The original document contains color images.	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** SAR
Number of Pages 7	

Figure 1 illustrates our methodology. EADSIM is run from a Unix script rather than the normal graphical user interface. We store the proper distributions for reliability, accuracy, and timelines in a database and random variates are drawn from these distributions offline between Monte Carlo trials. The Unix script inserts the variates in the proper EADSIM input files and automatically launches an EADSIM run. This process repeats until the desired number of Monte Carlo trials is achieved. Post-processing yields a distribution of protection effectiveness as a function of scenario time. We assess FoS confidence from that distribution.

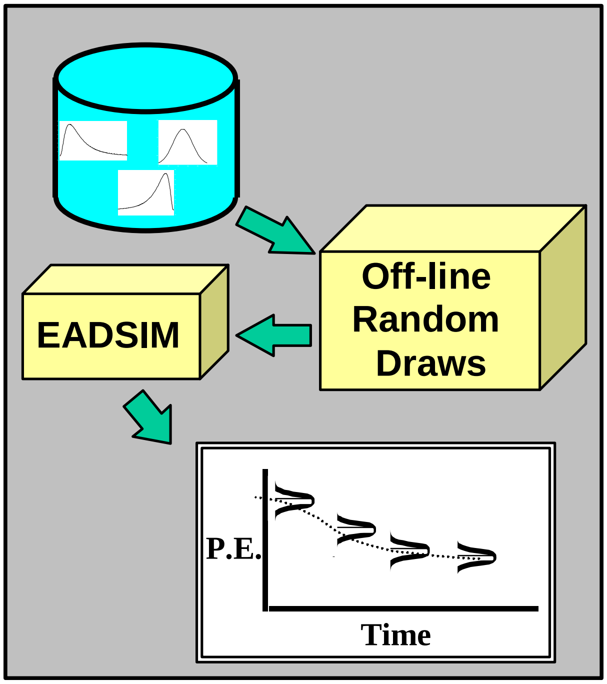

Figure 1. Schematic of the Methodology.

Workarounds are necessary to properly model some aspects of reliability, accuracy, and timelines in EADSIM. The reliability and availability of battlefield units like command centers and radars can only be modeled by scripting on/off times. Thus, our offline methodology randomly draws downtimes from Exponential Distributions and changes the appropriate on/off times for battlefield units.

Interceptor accuracy and endgame reliability are combined into a single Probability of Kill (Pk) in EADSIM, so our offline methodology draws separate random variates for interceptor accuracy and reliability from Binomial Distributions and combines them into the single Pk needed for EADSIM.

Timelines are easily modeled by drawing delay times from Lognormal Distributions and inserting them in the appropriate EADSIM input files.

TEST SCENARIO

To test the methodology we devised the simplified EADSIM scenario shown in Figure 2. A whimsical conflict between northern and southern Florida was chosen to keep the proof-of-principle unclassified and because of Florida's similarity to other theaters of interest.

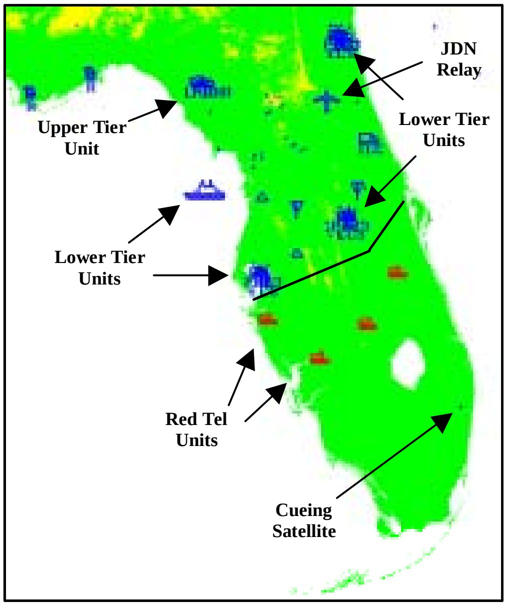

Figure 2. Scenario for the Proof-of-Principle.

South Florida launches one hundred eight Theater Ballistic Missiles (TBMs) at North Florida over four days of war. Thirty missiles are launched on the first day, followed by raids of forty, twenty, and eighteen missiles on the subsequent days. Approximately one-quarter of the TBMs are targeted against defensive units.

One ground-based upper tier system, one sea-based lower tier system, and three ground-based lower tier units defend North Florida. The ground-based lower tier units all have unlimited interceptor inventories, but the upper tier and the sea-based units are constrained to thirty and twenty-five interceptors each, respectively. A satellite is available to cue

defensive radars. The firing doctrine is shoot-look-shoot for the upper tier and salvo-two for the lower tier.

Table 1 lists the parameters we varied for the proof-of-principle. These are only a small subset of the parameters that may be needed to fully assess confidence in FoS performance. We selected the parameters in Table 1 because, taken together, they demonstrate all the workarounds necessary to conduct stochastic force-on-force experiments with EADSIM. Reliability, accuracy, and timeline parameters are all represented.

System	***Stochastic Parameter***
Threat TBM	Reliability Accuracy
GB Upper Tier	Launcher delay time Interceptor reliability Interceptor accuracy Radar reliability/availability Radar repair/downtime
GB Lower Tier	Launcher delay time Interceptor reliability Interceptor accuracy Radar reliability/availability Radar repair/downtime
SB Lower Tier	Launcher delay time Interceptor reliability Interceptor accuracy Radar reliability/availability Radar repair/downtime
Cueing System	Time for cue to reach radar

Table 1. Parameters that were Stochastically Varied in the Proof-of-Principle Test.

Figure 3 shows the distributions we selected for the experiment. For simplicity, the same distributions were applied to all the weapon systems. The distributions in Figure 3 are notional and were contrived for demonstration purposes only. Data collected in weapon system tests and BMDO FoS integration tests will be necessary to characterize performance. We believe Bayesian statistics will be necessary to develop these distributions from limited test data.

Cumulative FoS protection effectiveness (PE) was the primary measure of effectiveness for the proof-of-principle test. Cumulative PE is defined as the ratio of TBMs killed to the total number of incoming TBMs since the beginning of the scenario. Its value changes over the course of a scenario as the threat order of battle changes and as defensive weapon systems deplete inventory or change doctrine. Figure 4 illustrates PE for a single EADSIM Monte Carlo trial. The last value of PE is the most important because it represents FoS performance over the entire scenario.

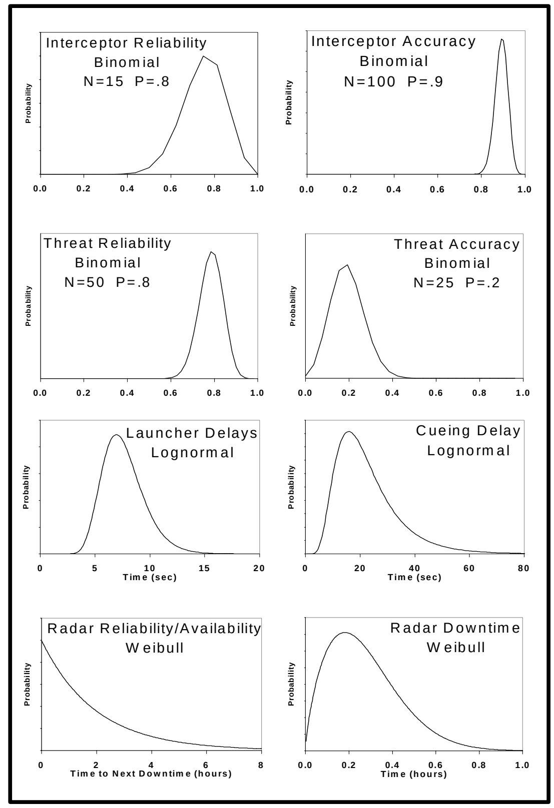

Figure 3. Input Distributions for the Test.

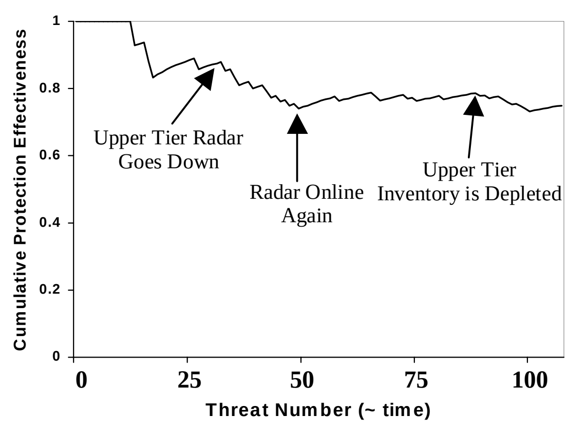

Figure 4. Cumulative PE for a Single Monte Carlo Trial.

RESULTS

We ran 250 Monte Carlo trials for the Florida Scenario, each with a unique set of random variates drawn from the distributions in Figure 3. FoS protection effectiveness for the 250 runs is shown in Figure 5.

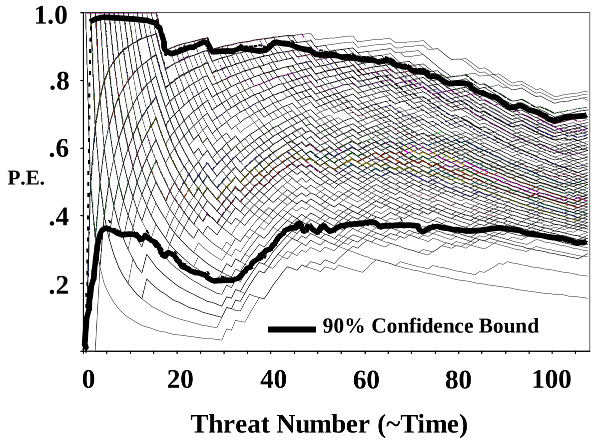

Figure 5. FoS Protection Effectiveness and Confidence for 250 Monte Carlo Trails.

As shown in the figure, FoS performance is highly unpredictable in the presence of the stochastic variations. After 30 TBMs FoS protection effectiveness ranges between 3% and 90%. At the end of the scenario the spread of FoS protection effectiveness is between 16% and 77%. The 90th percentile upper and lower confidence bounds on FoS protection effectiveness at the end of the scenario are 33% and 70%.

So what happened? Why is there so much variation? We first did a sanity check on the input distributions. Some people might argue our notional estimates of reliability/availability are too severe. The mean time to failure/downtime for all defense radars is approximately two hours. However, our estimate for radar repair/downtime is also low. A statistical combination of the two distributions (reliability and repair) results in a mean radar downtime for the radars of only 8% of the time.

The reality is that stochastic variations do matter. Testers and system evaluators must consider the adverse effects of random variations and uncertainties in weapon system performance. Expected value solutions are not adequate.

Figure 6 is another way of looking at protection effectiveness and reveals more insight into the results. It is a 3-dimensional histogram of FoS protection effectiveness as a function of threat number. The PE and Threat Number Axes lie in the horizontal plane, while the vertical axis is the number of occurrences of protection effectiveness in bins of .02-size.

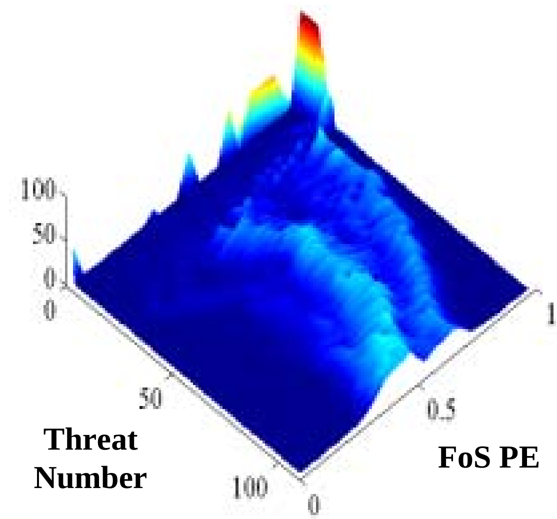

Figure 6. 3-D Histogram of FoS Protection Effectiveness for 250 Monte Carlo Trials.

Bi-modal behavior of the FoS is clearly evident in Figure 6. The two distributions correspond to FoS performance in two operating modes, single-tier defense and two-tier defense. The bi-modal behavior becomes less distinct near the end of the scenario because the upper tier unit in the Florida scenario has depleted its weapons by that time in most of the Monte Carlo trials.

Figure 7 is a horizontal cross-cut of the 3-dimensional protection effectiveness histogram at the end of the scenario, that is, at Threat Number 108. The bi-modal behavior is still evident.

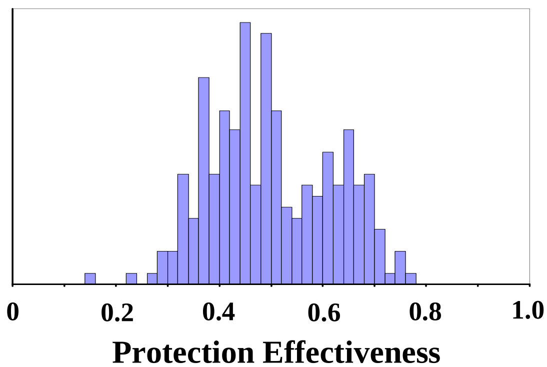

Figure 7. 2-Dimensional Histogram for FoS Protection Effectiveness at the End of the Scenario.

Sensitivity Analysis

Sensitivity analysis will be an essential ingredient of future FoS performance assessments. Our goal is to better characterize system behavior and to identify key performance drivers. There are hundreds of parameters in force-on-force simulations that could be stochastically varied. Hopefully, through the use of sensitivity analysis, these hundreds of parameters can be filtered to one or two dozen critical uncertainties. Once the drivers are known then parameters of lesser importance can be ignored in Monte Carlo analysis. Also, feedback from sensitivity analysis to the test community is essential so risk is efficiently identified and reduced through testing.

Factorial experiment designs are most often used for sensitivity analysis. In a factorial design high and low values are chosen for all the stochastic parameters and simulations are run for every possible combination of these factors. The main effects (i.e., the average change in system response due to a change in an individual factor) and interactions between the factors are computed from results of the simulations.

Ballistic missile defense entails too many stochastic parameters to efficiently conduct factorial sensitivity analysis. A factorial design requires 2^K simulation runs, where K is the number of stochastic parameters. Our simplistic proof-of-principle considered only eighteen factors (i.e., the factors listed in Table 1). A factorial sensitivity design would require:

$$2^{18} = 262{,}144 \text{ Simulation Runs}$$

The number of simulation runs increases exponentially with the number of factors. Obviously a more efficient method of estimating ballistic missile defense sensitivities is needed. Many candidate solutions to this dilemma are suggested in experimental design literature.

We selected the Random Balance Screening Methodology to conduct a trial sensitivity analysis on FoS performance as characterized by our notional distributions and force-on-force simulation. Random Balance Screening is a popular experimental design technique first introduced by Budne (1959).

An experiment design matrix is constructed with the parameters to be studied along the horizontal and the experimental conditions vertically. Each column is dealt an equal number of randomly distributed high and low (i.e., +90% confidence and -90% confidence) performance estimates. The number of experiments must be an even number so the number of +'s and –'s are balanced in each column. Increasing the number of experiments improves statistical confidence in the results.

We constructed a random balance sensitivity design for the BMDO FoS with 160 EADSIM experiments using the 90^{th} percentile high and low confidence bounds from the notional distributions in Figure 3. For simplicity, we combined the interceptor accuracy and reliability distributions into a single Pk distribution and the radar reliability/availability and repair/downtime distributions into a single distribution called radar fractional downtime (i.e., the percentage of time the radar is not operating). By combining these distributions for all three defensive weapon systems we reduced the number of stochastic parameters from eighteen to eleven.

We used a simple least squares estimator suggested by Mauro (1986) to compute the main effects from the 160 EADSIM runs:

$$\beta_j = \frac{(PE_{+j} - PE_{-j})}{2}$$

Where β_j is the main effect for the j^{th} factor, PE_{+j} is the average protection effectiveness of the 80 runs having the j^{th} factor at the high confidence bound, and PE_{-j} is the average protection effectiveness of the 80 runs having the j^{th} factor at the low confidence bound.

For a first test of the methodology we considered only main factor sensitivities at the end of the scenario. We did not consider non-linear interactions between the factors or sensitivities at other times in the scenario.

Figure 8 shows the main factor sensitivities for the proof-of-principle scenario.

The results are somewhat surprising. Their interpretation led to additional insights about the proof-of-principle scenario.

Most importantly, the Ground-Based (GB) Lower Tier Weapon System clearly dominates the outcome of FoS protection effectiveness for the Florida scenario. The three main effects with the largest magnitudes are, in order, the GB Lower Tier interceptor Pk, radar downtime, and launcher time delay. The GB Lower Tier System dominates FoS performance (we learned in hindsight) because our scenario has three GB Lower Tier Units, each having unlimited inventory, while there is only one Sea-Based (SB) Lower Tier Unit and one GB Upper Tier Unit, both with limited inventory. If the Upper Tier Unit were given unlimited inventory the sensitivity results would probably change drastically.

There were more surprises. We defined the experiment so all the main factor sensitivities would be positive, or so we thought. We arbitrarily arranged the upper and lower confidence bounds for each parameter so FoS protection effectiveness would always increase. For example, the high confidence bounds for the GB Lower Tier was defined as a $+j^{th}$ term, expecting that FoS PE would increase with a higher interceptor Pk. The high confidence bound for the Threat Pk was defined as a $-j^{th}$ term, expecting that FoS PE would *decrease* with a higher Threat Pk. As shown in Figure 8, our intuition was not always right.

FoS protection effectiveness increases as downtime for the SB Lower Tier Unit increases. This is contrary to common sense and forced us to take another look at the simulation results. We found a deconfliction problem between the SB Lower Tier and GB Upper Tier Units, causing them to simultaneously engage some of the same TBMs. The result is interceptor wastage and, with a limited inventory, the SB Lower Tier Unit cannot afford to waste interceptors. Better FoS protection effectiveness occurs when the SB Lower Tier Unit is down longer, increasing the likelihood that the GB Upper Tier Unit will expend its inventory while the SB Lower Tier is not operational. The SB Unit wastes fewer missiles when it begins to shoot again and FoS protection effectiveness improves.

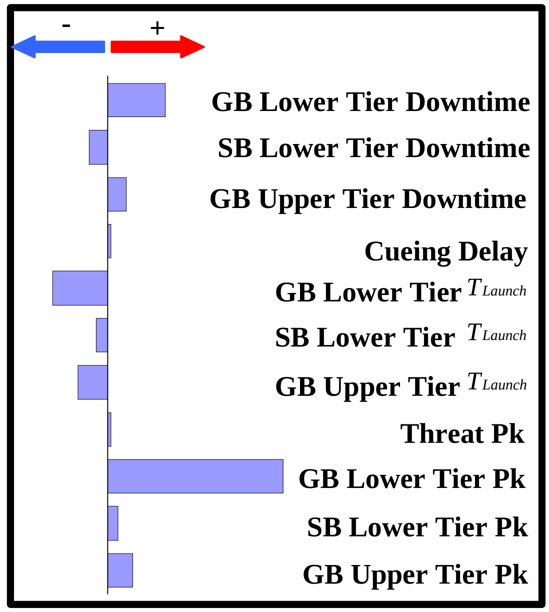

Figure 8. Results of Random Balance Sensitivity Analysis

The sensitivities to launcher delays are also negative in Figure 8 indicating that, contrary to expectations, FoS protection effectiveness increases as the launcher delays increase. We are still investigating why this occurred.

The sensitivities to cueing time and threat Pk are nearly zero. Cueing was not expected to be a significant factor in the Florida scenario because the threat ranges are too short for space-based cueing to be effective. Catastrophic FoS performance was observed in Monte Carlo trials whenever TBMs destroyed defensive batteries but, as the sensitivity analysis shows, TBM accuracy is apparently too low to statistically drive FoS performance. Sensitivity analysis is not useful for predicting the likelihood or effects of low probability events.

As the above discussion implies, FoS performance is often driven by scenario-specific factors. Projecting FoS performance with one scenario is not adequate. A variety of scenarios are needed to assess performance over the entire envelope of FoS missions and operating regimes.

CONCLUSION

We have demonstrated a methodology to extend BMDO test results to the theater-level using force-on-force simulation. The product of this methodology is quantified confidence in FoS performance and a better understanding of the underlying factors that influence performance.

The importance of stochastic variations is clearly evident in results from our first test of the methodology. Uncertainties in weapon system accuracy, timelines, and reliabilities can cause FoS protection effectiveness to deviate radically from expected value solutions. Testers and system evaluators need to identify these uncertainties, understand their effects, and reduce the potential risks to FoS performance as efficiently as possible. We believe our methodology will contribute to this process.

References

Budne, T. A. (1959). *The Application of Random Balance Designs*. Technometrics Vol. 3, pages 311-351.

Mauro, C. A. (1986). *Efficient Identification of Important Factors in Large Scale Simulations.* Proceedings of the 1986 Winter Simulation Conference.

Author Biographies

Barry Mitchell is a member of the Principal Staff at the Applied Physics Laboratory (APL). He has a B.S. in Aerospace Engineering from the University of Cincinnati and a M.S. in Applied Physics from the Johns Hopkins University. Since joining APL in 1986, Mr. Mitchell has been involved in test & evaluation and modeling & simulation projects related to the Trident Missile and ballistic missile defense programs. He previously worked for the McDonnell Douglas Corporation in Houston, Texas. E-mail: barry_mitchell@jhuapl.edu

Thomas Spriesterbach is a senior engineer at the Johns Hopkins Applied Physics Laboratory in Laurel, Maryland. He is a graduate of the United States Naval Academy with a Bachelors of Science Degree in Aerospace Engineering and a graduate of the Naval Postgraduate School with a Masters of Science Degree in Astronautical Engineering. Currently, he is working on threat trajectory modeling for ballistic missile defense programs and missile analysis of the Trident weapons system. E-mail: sprietp1@jhuapl.edu